Explanation of Correction Factor

The BSARs described in Tables 4-1 to 4-4 of the Bioaccumulation Modeling Report were derived using log-transformed data, and thus the mean sediment or tissue concentrations predicted using the BSARs are also in log-units (e.g., ln[mg/kg]). In order to back-transform the values, the predicted mean concentrations were exponentiated, returning predicted concentrations to the original scale. Although this back-transformation allows for the predictions to be returned to the original, and easier-to-comprehend scale, back-transformation introduces bias, such that the exponentiated concentration (in the original scale) is not necessarily an accurate prediction of the mean (Duan 1983).[1] In order to address this bias, a correction (or "smearing") factor can be incorporated into the estimation of the exponentiated mean. Duan (1983) describes the rationale and provides a method for calculating correction factors, and Helsel and Hirsch (2002) provide elaboration on the topic. Correction factors were calculated using the equation in Duan (1983), which was programmed with R software. The correction factor for a linear model is defined as the average value of exponentiated residuals (i.e., model predictions minus empirical data). The R script used to calculate smearing factors is provided in a separate text file called "smear.txt", and the data used to calculate correction factors (referenced in text file) is provided in a file called "ph_bsar.csv". Both files are necessary to reproduce the correction factor calculations.

BSAR equations were developed with the independent variable (Y) equal to the tissue concentration and the dependent variable (X) equal to the sediment concentration, as shown in the equation below.[2] In this equation , Y is shown to be transformed by the natural log.

$$X = EXP\left(\frac{(\ln(Y) - \ln(F) - a)}{b}\right)$$

Where:

Y = independent variable
X = dependent variable
a = model intercept
b = model slope
F = correction factor

This equation describes the BSAR model run in reverse, so that sediment concentrations are predicted for a given tissue concentration. PRG values can be calculated using this form of the BSAR equation by setting 'Y' (tissue concentration) equal to a target tissue level and solving for 'X' (sediment concentration). The smearing factor 'F' is included to correct for the bias introduced due to exponentiation (the 'EXP' term) of model output.

[1] Exponentiating the arithmetic mean of log-data provides an estimate of the geometric mean or median rather than the arithmetic mean (all on the original scale) (Helsel and Hirsch 2002)

[2] Also Equation 4-3 of the Bioaccumulation Modeling Report

References

Duan N. 1983. Smearing estimate: A nonparametric retransformation method. J Am Stat Assoc 78(383):605-610.

Helsel DR, Hirsch RM. 2002. Statistical methods in water resources. Chapter A3. Techniques of water-resources investigations of the United States Geological Survey, Book 4, Hydrologic analysis and interpretation. US Geological Survey, Washington, DC.